

**PERBANDINGAN SENSITIVITAS ANTARA METODE MANTEL-
HAENSZEL, REGRESI LOGISTIK DAN *IRT-LIKELIHOOD RATIO*
PADA PENDETEKSIAN *DIFFERENTIAL ITEM FUNCTIONING***



KERDID SIMBOLON
7817090934

Disertasi yang Ditulis untuk Memenuhi Sebagian Persyaratan
dalam Mendapatkan Gelar Doktor

PASCASARJANA
UNIVERSITAS NEGERI JAKARTA
2019

ABSTRAK

Penelitian ini bertujuan menguji perbedaan sensitivitas 3 metode deteksi *DIF* (*Differential Item Functioning*) yakni: *IRT-Likelihood Ratio* (*IRT-LR*), Mantel-Haenszel (MH) dan Regresi Logistik (RL), pada pendeteksian *DIF* Ujian Nasional (UN) Matematika SMA Program IPA Tahun 2015 di wilayah Sumatera Utara, Indonesia. Tepatnya bertujuan menentukan metode paling sensitif dalam mendeteksi *DIF* atau “mengendus” keberadaan *DIF* bahkan ketika derajat *DIF* kecil. Butir *DIF* adalah butir yang menguntungkan kelompok tertentu dan merugikan kelompok lain dengan kemampuan yang sama. *DIF* ditinjau dari dua jenis pengelompokan, *pertama* jenis kelamin (JK), dimana laki-laki sebagai kelompok referensi dan perempuan sebagai kelompok fokus; *kedua*, jenis sekolah (JS), dimana SMA Negeri sebagai kelompok referensi dan SMA Swasta sebagai kelompok fokus. Penelitian ini menggunakan metode eksperimen dengan disain 3x1, 1 faktor (metode) dengan 3 perlakuan (metode deteksi *DIF* yang berbeda). Populasi mencakup keseluruhan 5 jenis paket soal UN Matematika 2015 (kode 1107, 2207, 3307, 4407 dan 5507). Paket 2207 terambil sebagai sampel terdiri atas 5000 peserta (3067 perempuan, 1933 laki-laki; 2427 siswa SMA Swasta, 2573 siswa SMA Negeri), dengan 40 butir soal UN. Seleksi butir teori tes klasik (TTK) diterapkan terhadap 40 butir UN dan meloloskan 32 butir. Seleksi teori respon butir (TRB) meloloskan 18 butir yang memenuhi. Dengan program R 3.333 dan *IRTLRDIF 2.0* pada kelompok JK didapat 5 butir *DIF* oleh *IRT-LR*, 4 butir oleh RL, dan 3 butir oleh MH. Pada kelompok JS didapat 18 butir *DIF* oleh *IRT-LR*, 32 butir oleh RL, dan 25 butir oleh MH. Dalam uji sensitivitas metode *DIF*, dibentuk 6 kelompok analisis: (4400,40), (4400,32), (4400,18), (3000,40), (3000,32), (3000,18), serta dibangkitkan 40 set data acak dan dilakukan 40 kali deteksi *DIF* pada setiap kelompok. Dengan data tidak *fit* model, model logistik 3 parameter (3PL) dipilih sebagai model paling sesuai. Hasil uji Tukey menunjukkan metode *IRT-LR* lebih sensitif dari MH dan RL dalam kelompok (4400,40) dan (3000,40). *IRT-LR* teruji lebih sensitif dari MH pada kelompok (4400,32) dan (3000,32) tetapi tidak lebih sensitif dari RL. Kedua metode MH dan RL berbalik menjadi lebih sensitif dari *IRT-LR* dalam kelompok (4400,18) dan (3000,18). Metode RL konsisten teruji lebih sensitif dari MH pada seluruh kelompok analisis. Sementara pada kelompok JS, seluruh hipotesis penelitian teruji signifikan dimana RL dan MH berbalik menjadi lebih sensitif dari *IRT-LR*, dan metode RL konsisten lebih sensitif dari MH.

Kata Kunci: Ujian Nasional, *DIF*, *IRT-Likelihood Ratio*, Mantel-Haenszel, Regresi Logistik

ABSTRACT

The purpose of this study was to examine the differences in sensitivity of 3 DIF (Differential Item Functioning) detection methods, namely: IRT-Likelihood Ratio (IRT-LR), Mantel-Haenszel (MH) and Logistic Regression (RL), in detecting DIF Mathematics National Examination (UN) academic year of 2015 in North Sumatra, Indonesia. Precisely to determine the most sensitive method of detecting DIF or "sniffing" the presence of DIF even when the DIF degree is small. DIF item is item that benefit certain group participants and disadvantage other groups with similar abilities. The presence of DIF is reviewed on two types of grouping of participants. First by gender (JK) where men as a reference group (R) and women as focus group (F); second by type of school (JS), where public high school as a reference group (R) and private high school as focus group (F). This study uses an experimental method with a 3x1 design, with 1 factor (method) with 3 treatments (different DIF detection methods). The population includes all 5 types of 2015 Mathematics UN question packages (codes: 1107, 2207, 3307, 4407 and 5507). The 2207 package was taken as a sample consisting of 5000 participants (3067 women, 1933 men; 2427 Private High School students, 2573 Public High School students), with 40 UN items. Classic test theory (TTK) item selection was applied to 40 UN items and passed 32 items. Item response theory selection (TRB) passed 18 items that fulfilled. With the R 3.333 program and IRTLRDIF 2.0, in the JK group, 5 items were detected as DIF by IRT-LR, 4 items by RL, and 3 items by MH. In the JS group, there were 18 items were detected as DIF by IRT-LR, 32 items by RL, and 25 items by MH. In the sensitivity test of the DIF methods, 6 analysis groups were formed: (4400,40), (4400,32), (4400,18), (3000,40), (3000,32), (3000,18), 40 data sets were generated and 40 times DIF detection was performed in each group. With data not fit models, a three parameter logistic model (3PL) is chosen as the most suitable model. In the JK group, Tukey's test results showed the IRT-LR method was more sensitive than MH and RL in the groups (4400,40) and (3000,40). IRT-LR was tested to be more sensitive than MH in the groups (4400,32) and (3000,32) but not more sensitive than RL. Both MH and RL methods turn out to be more sensitive than IRT-LR in groups (4400,18) and (3000,18). The RL method was consistently tested to be more sensitive than MH in the entire analysis group. While in the JS group, all research hypotheses were tested as significant where RL and MH turned to be more sensitive than IRT-LR, and the RL method was consistently more sensitive than MH.

Key words: National Examination, DIF, IRT Likelihood Ratio, Mantel-Haenszel, Logistic Regression

**PERSETUJUAN DEWAN PENGUJI
DIPERSYARATKAN UNTUK UJIAN TERBUKA /PROMOSI DOKTOR**

Promotor,

Ko-Promotor,



Prof. Dr. Yetti Supriyati, M.Pd.



Prof. Dr. Ir. Dali Santun Naga, MMSI.

Tanggal: 24-10-2019

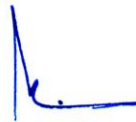
Tanggal: 24-10-2019

NAMA

TANDA TANGAN

TANGGAL

Prof. Dr. Ilza Mayuni, M.A.
(Ketua)¹



Dr. Wardani Rahayu, M.Si.
(Sekretaris)²



Nama : Kerdid Simbolon

No. Registrasi : 7817090934

Program Studi : Penelitian dan Evaluasi Pendidikan

Tanggal Lulus :

¹ Plt. Direktur Pascasarjana Universitas Negeri Jakarta

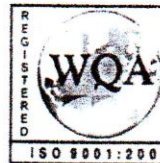
² Koordinator Program Studi Penelitian dan Evaluasi Pendidikan



*Building
Future
Leaders*

**KEMENTERIAN RISET, TEKNOLOGI DAN PENDIDIKAN TINGGI
UNIVERSITAS NEGERI JAKARTA
PASCASARJANA**

Kampus Universitas Negeri Jakarta, Jl. Rawamangun Muka Jakarta-Timur 13220
Telp. (021) 4721340, Fax (021) 4897047, website: <http://pps.unj.ac.id>, e-mail: tu.pps@unj.ac.id



Cert.No.QS41

LEMBAR PERNYATAAN

Saya menyatakan dengan sesungguhnya bahwa Disertasi yang saya susun sebagai syarat untuk memperoleh gelar Doktor dari Program Pascasarjana Universitas Negeri Jakarta seluruhnya merupakan hasil karya saya sendiri.

Adapun bagian-bagian tertentu dalam penulisan Disertasi yang saya kutip dari hasil karya orang lain telah dituliskan sumbernya secara jelas sesuai dengan norma, kaidah dan etika penulisan ilmiah.

Apabila di kemudian hari ditemukan seluruh atau sebagian Disertasi ini bukan hasil karya saya sendiri atau adanya plagiat dalam bagian-bagian tertentu, saya bersedia menerima sanksi pencabutan gelar akademik yang saya sandang dan sanksi-sanksi lainnya sesuai peraturan perundangan yang berlaku.

Jakarta, 2 Agustus 2019



[Signature]
Kerdid Simbolon



KEMENTERIAN PENDIDIKAN DAN KEBUDAYAAN
UNIVERSITAS NEGERI JAKARTA
UPT PERPUSTAKAAN

Jalan Rawamangun Muka Jakarta 13220
Telepon/Faksimili: 021-4894221
Laman: lib.unj.ac.id

**LEMBAR PERNYATAAN PERSETUJUAN PUBLIKASI
KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIS**

Sebagai sivitas akademika Universitas Negeri Jakarta, yang bertanda tangan di bawah ini, saya:

Nama : Kerdid Simbolon
NIM : 7817090934
Fakultas/Prodi : Pascasarjana / Penelitian dan Evaluasi Pendidikan
Alamat email : kerdidsimbolon@gmail.com

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada UPT Perpustakaan Universitas Negeri Jakarta, Hak Bebas Royalti Non-Eksklusif atas karya ilmiah:

☐ Skripsi ☐ Tesis ☒ Disertasi ☐ Lain-lain (.....)

yang berjudul :

Perbandingan Sensitivitas Antara Metode Mantel-Haenszel, Regresi Logistik dan *Irt-likelihood*
Ratio Pada Pendeteksian *Differential Item Functioning*

Dengan Hak Bebas Royalti Non-Eksklusif ini UPT Perpustakaan Universitas Negeri Jakarta berhak menyimpan, mengalihmediakan, mengelolanya dalam bentuk pangkalan data (*database*), mendistribusikannya, dan menampilkan/mempublikasikannya di internet atau media lain secara *fulltext* untuk kepentingan akademis tanpa perlu meminta ijin dari saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan atau penerbit yang bersangkutan.

Saya bersedia untuk menanggung secara pribadi, tanpa melibatkan pihak Perpustakaan Universitas Negeri Jakarta, segala bentuk tuntutan hukum yang timbul atas pelanggaran Hak Cipta dalam karya ilmiah saya ini.

Demikian pernyataan ini saya buat dengan sebenarnya.

Jakarta, 23 April 2020

Penulis

(Kerdid Simbolon)

KATA PENGANTAR

Terpujilah Tuhan Yesus Kristus atas pertolongan-Nya yang ajaib, sehingga penulis dapat menyelesaikan disertasi yang berjudul: **Perbandingan Sensitivitas antara Metode *IRT-Likelihood Ratio*, Mantel-Haenszel dan Regresi Logistik pada Pendeteksian *Differential Item Functioning (DIF)***. Adapun disertasi ini disusun dalam rangka memenuhi sebagian persyaratan untuk memperoleh gelar Doktor Penelitian dan Evaluasi Pendidikan pada Program Pascasarjana Universitas Negeri Jakarta (UNJ).

Penulis mengakui bahwa disertasi ini dapat diselesaikan berkat dukungan dan bantuan dari banyak pihak. Oleh karena itu, penulis mengucapkan terimakasih sebesar-besarnya kepada semua pihak yang secara langsung maupun tidak langsung telah memberikan kontribusi yang berharga. Secara khusus penulis menyampaikan terimakasih kepada Prof. Dr. Yetty Supriyati, M.Pd. dan Prof. Dr. Dali Santun Naga sebagai Promotor dan Co-Promotor yang telah membimbing dan mengarahkan penulis selama penyusunan disertasi ini dari awal hingga selesai. Penulis juga berterimakasih kepada Dr. Wardani Rahayu, M.Si. sebagai Ketua Program Studi Penelitian dan Evaluasi Pendidikan Program Pascasarjana Universitas Negeri Jakarta (UNJ) yang telah bersedia memberi masukan berharga dalam penulisan disertasi ini.

Disertasi ini tentu tidak terlepas dari berbagai kekurangan di sana-sini, namun demikian harus tetap diakui bahwa karya mulia ini merupakan hasil olah pikir, kerja keras dan kesabaran.

Jakarta, 26 Agustus 2019

Penulis,



Kerdid Simbolon

DAFTAR ISI

	<i>Halaman</i>
Kata Pengantar.....	i
Abstrak	ii
<i>Abstract</i>	iii
Daftar Isi.....	iv-vii
Daftar Tabel.....	viii-ix
Daftar Gambar	x
Daftar Lampiran	xiii-xiv
BAB I PENDAHULUAN	1
1.1. Latar Belakang Masalah	1
1.2. Identifikasi Masalah	17
1.3. Pembatasan Masalah	20
1.4. Perumusan Masalah	21
1.5. Tujuan Penelitian	22
1.6. Signifikansi Penelitian	22
1.7. Kebaruan Penelitian (<i>State of the Art</i>)	24
BAB II KAJIAN TEORETIK	26
2.1. Kajian Pustaka	26
2.1.1. Pengukuran Dalam Pendidikan	26
2.1.1.1. Analisis Butir Soal	29
2.1.2. Pengukuran Menurut Teori Tes Klasik	34
2.1.3. Pengukuran Menurut Teori Respon Butir	42
2.1.3.1. Unidimensi.....	45
2.1.3.2. Independensi Lokal	45
2.1.3.3. Ivariansi Parameter	47
2.1.3.4. Pemilihan Model.....	49
2.1.4. Bias, Dampak Butir & <i>Differential Item Functioning (DIF)</i> .	58
2.1.5. <i>Differential Item Functioning</i>	67
2.1.6. Tipe <i>Differential Item Functioning</i>	70
2.1.7. Fungsi Informasi.....	74
2.1.8. Metode Deteksi <i>Differential Item Functioning</i>	78
2.1.8.1. Metode deteksi DIF Parametrik	79
2.1.8.2. Metode deteksi DIF Nonparametrik	83
2.1.8.3. Metode Mantel-Haenszel	94
2.1.8.4. Metode Regresi Logistik	101
2.1.8.5. Metode <i>IRT-Likelihood Ratio</i>	109

2.2.	Hasil-hasil Penelitian Relevan	114
2.3.	Kerangka Berpikir.....	144
2.3.1.	Perbedaan sensitivitas pendeteksian <i>Differential Item Functioning</i> menggunakan metode Mantel-Haenszel dan metode <i>IRT-Likelihood Ratio</i>	145
2.3.2.	Perbedaan sensitivitas pendeteksian <i>Differential Item Functioning</i> menggunakan metode Regresi Logistik dan metode <i>IRT-Likelihood Ratio</i>	149
2.3.3.	Perbedaan sensitivitas pendeteksian <i>Differential Item Functioning</i> menggunakan metode Mantel-Haenszel dan metode Regresi Logistik	152
2.4.	Pengajuan Hipotesis Penelitian.....	155
BAB III	METODOLOGI PENELITIAN	156
3.1.	Tempat dan Waktu Penelitian.....	156
3.2.	Metode Penelitian	156
3.3.	Sumber Data	160
3.4.	Populasi dan Teknik Pengambilan Sampel	161
3.5.	Validitas dan Reliabilitas Instrumen Tes.....	164
3.6.	Bagan Alur Kegiatan Penelitian	165
3.7.	Prosedur Penelitian	166
3.7.1.	Seleksi Butir Soal	167
3.7.2.	Seleksi Model (Pencocokan Model)	169
3.8.	Teknik Analisis Data.....	170
3.8.1.	Uji Normalitas	172
3.8.2.	Uji Kehomogenan Varians	173
3.8.3.	Pengujian Hipotesis	173
3.9.	Hipotesis Statistik	175
BAB IV	HASIL PENELITIAN DAN PEMBAHASAN	177
4.1.	Deskripsi Data.....	177
4.1.1.	Data Awal Penelitian	177
4.1.2.	Statistik dan Distribusi Data.....	180
4.1.3.	Seleksi Butir Menurut Teori Tes Klasik	185
4.2.	Analisis Butir dan Penerapan Model Teori Respon Butir	188
4.2.1.	Pembersihan Data	188
4.2.2.	Seleksi Model	190
4.2.3.	Fungsi Informasi Tes	195
4.2.4.	Seleksi Butir Menurut Teori Respon Butir	199
4.2.5.	Parameter Butir UN Menurut Model Logistik 3PL	203
4.2.6.	Kurva Karakteristik Butir Model Logistik 3PL	206
4.2.7.	Fungsi Informasi Butir	208

4.2.8.	Uji Unidimensi	212
4.2.9.	Uji Independensi Lokal	214
4.2.10.	Uji Invariansi Parameter.....	214
4.2.11.	Parameter Kemampuan Peserta Ujian Nasional.....	216
4.2.12.	Pola Jawaban Ujian Nasional Matematika Tahun 2015.....	218
4.3.	Pendeteksian <i>DIF</i> Menggunakan Metode Mantel-Haenszel, Regresi Logistik, <i>IRT-Likelihood Ratio</i>	222
4.3.1.	Deteksi <i>DIF</i> (Kelompok Jenis Kelamin)	223
4.3.1.1.	Butir <i>DIF</i> Hasil Deteksi Metode Mantel-Haenszel.....	225
4.3.1.2.	Butir <i>DIF</i> Hasil Deteksi Metode Regresi Logistik.....	226
4.3.1.3.	Butir <i>DIF</i> Hsl. Deteksi Metode <i>IRT-Likelihood Ratio</i>	228
4.3.2.	Deteksi <i>DIF</i> (Kelompok Jenis Sekolah).....	231
4.3.2.1.	Butir <i>DIF</i> Hasil Deteksi Metode Mantel-Haenszel.....	234
4.3.2.2.	Butir <i>DIF</i> Hasil Deteksi Metode Regresi Logistik.....	236
4.3.2.3.	Butir <i>DIF</i> Hsl. Deteksi Metode <i>IRT-Likelihood Ratio</i>	238
4.3.2.4.	Perbandingan Jumlah Butir <i>DIF</i> Kel. JK dan JS.....	238
4.4.	Analisis Sensitivitas Metode <i>DIF</i> : Mantel-Haenszel, Regresi Logistik, <i>IRT-Likelihood Ratio</i> (40x iterasi).....	240
4.4.1.	Analisis Sensitivitas Metode <i>DIF</i> (Kel. Jenis Kelamin).....	245
4.4.1.1.	Deskripsi Butir <i>DIF</i> (Kelompok JK)	245
4.4.1.2.	Butir <i>DIF</i> Bersama (Kelompok JK).....	250
4.4.2.	Analisis Sensitivitas Metode <i>DIF</i> (Kel. Jenis Sekolah)	255
4.4.2.1.	Deskripsi Butir <i>DIF</i> (Kelompok JS)	255
4.4.2.2.	Butir <i>DIF</i> Bersama (Kelompok JS).....	260
4.5.	Uji Hipotesis Perbedaan Sensitivitas Metode <i>DIF</i>	265
4.5.1.	Uji Hipotesis Perbedaan Sensitivitas Metode <i>DIF</i> pada Kelompok Jenis Kelamin	265
4.5.1.1.	Uji Normalitas.....	265
4.5.1.2.	Uji Homogenitas	266
4.5.1.3.	Uji Hipotesis (Kelompok Analisis (4400,40))	266
4.5.1.4.	Uji Hipotesis (Kelompok Analisis (4400,32)).....	268
4.5.1.5.	Uji Hipotesis (Kelompok Analisis (4400,18))	270
4.5.1.6.	Uji Hipotesis (Kelompok Analisis (3000,40))	272
4.5.1.7.	Uji Hipotesis (Kelompok Analisis (3000,32))	274
4.5.1.8.	Uji Hipotesis (Kelompok Analisis (3000,18))	276
4.5.2.	Uji Hipotesis Perbedaan Sensitivitas Metode <i>DIF</i> pada Kelompok Jenis Sekolah	281
4.5.2.1.	Uji Normalitas.....	281
4.5.2.2.	Uji Homogenitas	281

4.5.2.3.	Uji Hipotesis (Kelompok Analisis (4400,40))	282
4.5.2.4.	Uji Hipotesis (Kelompok Analisis (4400,32))	284
4.5.2.5.	Uji Hipotesis (Kelompok Analisis (4400,18))	286
4.5.2.6.	Uji Hipotesis (Kelompok Analisis (3000,40))	288
4.5.2.7.	Uji Hipotesis (Kelompok Analisis (3000,32))	290
4.5.2.8.	Uji Hipotesis (Kelompok Analisis (3000,18))	292
4.6.	Pembahasan Hasil Penelitian	296
4.6.1.	Temuan pada Data Penelitian	297
4.6.2.	Pemenuhan asumsi TRB yang lemah	297
4.6.3.	Ketidakwajaran Pola jawaban UN	298
4.6.4.	Pemabahasan Hasil Uji Hipotesis Penelitian	300
4.6.4.1.	Pembahasan hasil penelitian (Kel. JK)	300
4.6.4.2.	Pembahasan hasil penelitian (Kel. JS)	307
4.7.	Keterbatasan Hasil Penelitian	313
BAB V	KESIMPULAN, IMPLIKASI DAN SARAN	315
5.1.	Kesimpulan	315
5.1.1.	Kesimpulan (Kelompok Jenis Kelamin)	315
5.1.2.	Kesimpulan (Kelompok Jenis Sekolah)	318
5.2.	Implikasi	320
5.3.	Saran	322
	DAFTAR PUSTAKA	327
	LAMPIRAN	333
	RIWAYAT HIDUP	460